

## Title

### Climatologies at high resolution for the Earth land surface areas

## Authors

Dirk Nikolaus Karger<sup>1</sup>, Olaf Conrad<sup>2</sup>, Jürgen Böhrer<sup>2</sup>, Tobias Kawohl<sup>2</sup>, Holger Kreft<sup>3</sup>, Rodrigo Wilber Soria-Auza<sup>3</sup>, Niklaus Zimmermann<sup>4</sup>, H. Peter Linder<sup>1</sup>, Michael Kessler<sup>1</sup>

## Affiliations

1. Department of Systematic and Evolutionary Botany, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland.

2. Institute of Geography, University of Hamburg, Bundesstrasse 55, 20146 Hamburg, Germany.

3. Biodiversity, Macroecology & Conservation Biogeography Group, University of Göttingen, 37077 Göttingen, Germany

4. Swiss Federal Research Institute WSL, Zürcherstr. 111, 8903 Birmensdorf, Switzerland.

corresponding author: Dirk Nikolaus Karger ([dirk.karger@systbot.uzh.ch](mailto:dirk.karger@systbot.uzh.ch)), Tel.: +41 44 634 8430.

## Abstract

High resolution information of climatic conditions is essential to many application in environmental sciences. Here we present the CHELSA algorithm to downscale temperature and precipitation estimates from the European Centre for Medium-Range Weather Forecast (ECMWF) climatic reanalysis interim (ERA-Interim) to a high resolution of 30 arc sec. The algorithm for temperature is based on a statistical downscaling of atmospheric temperature from the ERA-Interim climatic reanalysis. The precipitation algorithm incorporates orographic predictors such as wind fields, valley exposition, and boundary layer height, and a bias correction using Global Precipitation Climatology Center (GPCC) gridded and Global Historical Climate Network (GHCN) station data. The resulting data consist of a monthly temperature and precipitation climatology for the years 1979-2013. We present a comparison of data derived from the CHELSA algorithm with two other high resolution gridded products with overlapping temporal resolution (Tropical Rain Measuring Mission (TRMM) for precipitation, Moderate Resolution Imaging Spectroradiometer (MODIS) for temperature) and station data from the Global Historical Climate Network (GHCN). We show that the climatological data from CHELSA has a similar accuracy to other products for temperature, but that the predictions of orographic precipitation patterns are both better and at a high spatial resolution.

## Background & Summary

High resolution climate data are essential to many applications in environmental sciences. While many studies in these fields are conducted at a resolutions of  $\sim 1\text{km}^2$ , state of the art climate reanalyses often only represent climatic variations at spatial resolutions of  $0.5^\circ - 1^\circ$  (ca. 25 – 100 km at the equator) at a global scale. The gap between these spatial scales is often regionally bridged using satellite data<sup>1-3</sup>, via statistical downscaling<sup>4-7</sup>, or interpolation methods<sup>8</sup>, but climatologies based on statistical downscaling are not currently available on a global scale due to numerous methodological challenges<sup>9</sup>. While interpolated datasets are available<sup>8</sup>, they often fail to accurately predict certain factors such as precipitation in highly variable terrain<sup>10</sup>. To achieve a finer resolution, interpolation and regression techniques either use data from local climate observations such as the Global Historical Climate Network (GHCN)<sup>11,12</sup> and combine them with atmospheric predictors from gridded climate reanalyses

from the Global Precipitation Climatology Center (GPCC)<sup>13,14</sup>, Climate Research Unit (CRU)<sup>15,16</sup>, National Center for Atmospheric Research/National Center for Atmospheric Research (NCEP/NCAR)<sup>17</sup>, and the European Centre for Medium-Range Weather Forecast (ECMWF) climatic reanalysis interim (ERA-Interim)<sup>18</sup>.

Different interpolation techniques such as linear or inverse distance interpolations and geo-statistical kriging approaches have been used to delineate spatial high-resolution climatic information from local observations. These approaches allow for the inclusion of additional statistical parameters such as the standard error of an estimated value for assessing the statistical precision of spatial estimates<sup>19,20</sup>. However, although interpolation approaches such as the universal (regression) kriging<sup>21</sup> allow the integration of controlling land-surface parameters such as elevation, slope or aspect, satisfactory results still require a more or less regular distribution of input data and a proper representation of topo-climatic settings<sup>10</sup>. This is confounded by the very biased distribution of weather stations, which are mostly found in more accessible environments. This leads to a poor representation of climatic variability in mountainous regions or areas with intact lowland rainforest, such as the Amazon or Congo basin, leading to a poor representation of these areas in global climate surfaces<sup>8,10</sup>.

Statistical downscaling, on the other hand, exploits the observed relationship between large-scale circulation models (GCM) and local weather stations or tries to find transfer functions between them<sup>6,7,22</sup>. Using multivariate statistical analyses (e.g. product-moment or canonical correlation analyses) it is possible to identify sets of large-scale GCM variables and obtain empirical functions (e.g. regression equations) which can predict the local weather variations at the spatial scale of interest. Such statistical downscaling approaches are, however, also heavily dependent on the data distribution and generally function better in areas with high data density<sup>10</sup>. Especially on a global scale, statistical downscaling becomes problematic<sup>23</sup>, as the spatial distribution of weather stations changes through time. While measurements for a given predictor might be available in a certain month, they might be absent in another, leading to a generally high heterogeneity of the underlying climate records when time series of precipitation need to be calculated. While this does not affect static predictors such as elevation, slope, or aspect, statistical downscaling becomes especially problematic when highly dynamic predictors such as wind fields need to be integrated.

This general heterogeneity in the temporal and spatial distribution of such dynamic factors can lead to spurious correlations in specific months or specific regions, which can severely influence the parameters of a regression model. When specific predictors, such as windward or leeward sites of a mountain<sup>22,24</sup> change over the course of the year, the location of the climatic records does not change in accordance. Therefore, regression-based downscaling might, for example, detect a significant negative relationship between a station in the windward site of a mountain during one month, and a positive one in another, although atmospheric physics would always predict a positive relationship. Due to this problem, statistical downscaling and interpolation methods have often been applied to single regions<sup>25</sup>, while a global model is lacking.

To overcome the problem of heterogeneous spatial and temporal distribution of station data, and the problems encountered with spatial interpolation and regression approaches, we use a mechanistical downscaling algorithm for temperature and precipitation data provided from the ERA-Interim reanalysis<sup>18</sup> in combination with gauge derived products from the GPCC<sup>13</sup> and the GHCN<sup>11,12</sup> datasets. Applying this algorithm to a monthly dataset ranging from Jan. 1979 to Dec. 2013, we created a dataset with high spatial and temporal accuracy in topographic complex terrain globally, which might prove valuable in varied scientific applications that rely on high resolution climatic data.

## Methods

### Calculation of monthly temperature and precipitation values

ERA-Interim (developed at the European Centre for Medium-Range Weather Forecast, ECMWF), simulates 6-hourly large-scale atmospheric fields for 60 pressure levels between 1000 and 1 hPa globally with a horizontal resolution of 0.75 lat/long (T255)<sup>18,26,27</sup>. Since the ERA-Interim reanalysis combines modelling results with ground and radiosonde observations as well as remote sensing data using a data assimilation system, the free-atmospheric and surface fields can be considered as the best approximation of the current largescale atmospheric situation for every time step. Several studies reveal that ERA-Interim adequately captures the variability of relevant free-air meteorological parameters, even over complex terrain<sup>28–30</sup>.

## Temperature

Spatial variations of temperature is to a large degree determined by the vertical state of the troposphere and thus, if not affected by inversion layers, decrease with altitude<sup>31,32</sup>. The long term mean hypsometric temperature gradient covered in the ERA interim data accurately reflects the vertical distribution of moist- or dry-adiabatic lapse rates<sup>24</sup>. Typical temperature laps rates are in the order of  $-0.4$  to  $-0.8$  K/100 m with a characteristic seasonality. The corresponding temperature distribution pattern<sup>33</sup> is therefore closely related to the surface elevation<sup>22</sup>.

For our downscaling approach, we used the monthly means of daily means for temperature. Temperature lapse rates were calculated from the ERA Interim for pressure levels from 1000 hPa to 300 hPa, using linear regression for each grid ERA Interim cell and then reduced to sea level. Sea level was then interpolated between grid cells, and then projected back on the elevational surface of the DEM using the formula:

$$t_{elev} = \Gamma_d * elev + t_0 \quad [1]$$

where  $t_{elev}$  equal the Temperature at a given elevation,  $\Gamma_d$  equals the lapse rate,  $elev$  equals elevation at CHELSA resolution from GMTED2010<sup>34</sup>, and  $t_0$  equals the interpolated temperature at sea level.

Although temperature lapse rates are relatively linear, inversion layers or slight deviations in the linearity of the observed lapse rate can cause some over- or underestimation of  $t_0$  when the lapse rate is approximated from 1000 hPa to 300 hPa pressure levels especially in flat terrain. To correct for the possible deviations in temperatures from the mean, we used:

$$t_{elev\_cor} = t_{elev} + t_{era} - \overline{t_{elev}} \quad [2]$$

where  $t_{elev\_cor}$  is the corrected temperature,  $t_{era}$  is the mean daily temperature at ERA-Interim resolution, and  $\overline{t_{elev}}$  being the mean  $t_{elev}$  at ERA-Interim resolution.

## Precipitation

Globally, elevation is one of the main topo-climatic drivers of vertical precipitation gradients<sup>22,35–40</sup>. In the convective regimes of the tropics, precipitation amounts commonly increase up to the condensation level at about 1000-1500 m above the ground surface while the exponentially decreasing air moisture content in the mid- to upper troposphere results in a corresponding drying above the condensation level of tropical convection cluster systems (non linear precipitation lapse rates). Likewise, negative lapse rates typically occur in the extreme dry polar climates. At mid latitudes and in the subtropics, the frequent or even prevalent high reaching advection of moisture bearing air leads to increasing precipitation amounts of high mountain ranges such as the Alps<sup>41</sup> (linear precipitation lapse rates<sup>42</sup>). The reduced precipitation amounts at lower settings are firstly due to the transpiration of rain drops when falling through non-saturated, lower-air levels. Secondly, the vertical precipitation gradient in high mountain

ranges is often strengthened owing to the diurnal formation of autochthonous upslope breezes, which intensify cloud and precipitation formation in upper slope positions whilst the subsiding branch of these autochthonous local circulation systems along the valley axis leads to cloud dissolution and a corresponding reduction of precipitation rates in the valley bottoms. We approximated such orographic precipitation effects and took them into account in the CHELSA precipitation algorithm (Fig .1) as explained below.

### Wind effect correction

Orographic precipitation patterns<sup>43</sup>, caused by the uplift of moist air currents at the windward side of a mountain range and the intimately related rain shadow effect at leeward settings induced by the blockage of moisture-bearing air, are most common effects influencing small scale precipitation patterns<sup>41,43–46</sup>. Based on the assumption that the windward impact on the precipitation intensity depends on the prevailing wind direction at any given elevation of an orographic barrier, we used a wind index<sup>22,24</sup> to account for the expected higher precipitation at the windward sites of an orographic barrier.

We used u-wind and v-wind components at the 10 m level of ERA-Interim as underlying wind components. These two wind components were interpolated to the CHELSA grid resolution using a B-spline interpolation. As the calculation of a windward leeward index (hereafter: wind effect) requires a projected coordinate system, both wind components were projected to a world Mercator projection and then combined to a directional grid. The wind effect  $H$  with windward component  $H_W$  and the leeward component  $H_L$  were then calculated using:

$$H_W = \frac{\sum_{i=1}^n \frac{1}{d_{WHi}} \tan^{-1}\left(\frac{d_{WZi}}{d_{WHi}}\right)}{\sum_{i=1}^n \frac{1}{d_{LHi}}} + \frac{\sum_{i=1}^n \frac{1}{d_{LHi}} \tan^{-1}\left(\frac{d_{LZi}}{d_{LHi}}\right)}{\sum_{i=1}^n \frac{1}{d_{LHi}}} \quad [3]$$

$$H_L = \frac{\sum_{i=1}^n \frac{1}{d_{WHi}} \tan^{-1}\left(\frac{d_{LZi}}{d_{WHi}}\right)}{\sum_{i=1}^n \frac{1}{d_{LHi}}} \quad [4]$$

where  $d_{WHi}$  and  $d_{LHi}$  refer to the horizontal distances in windward and leeward direction and  $d_{WZi}$  and  $d_{LZi}$  are the corresponding vertical distances compared with the considered raster cell. The second summand in Eq. [3] accounts for the leeward impact of previously traversed mountain chains. The logarithmized horizontal distances in Eq. [4] lead to a longer-distance impact of leeward rain shadow. The final wind-effect parameter, which is assumed to be related to the interaction of the large-scale wind field and the local-scale precipitation characteristics, is calculated as  $H = H_L \times H_W$  and generally takes values between 0.7 for leeward and 1.3 for windward positions<sup>22</sup>. Eq. [3] and Eq. [4] were applied to each grid cell at the CHELSA resolution in a world Mercator projection. We used a corresponding search distance of 15 km for the wind effect, so that the algorithm follows a wind direction vector for 15 km to detect orographic barriers.

### Valley exposition correction

Although the wind effect algorithm can distinguish between the windward and leeward sites of an orographic barrier, it cannot distinguish extremely isolated valleys in high mountain areas. Such dry valleys are situated in areas where the wet air masses that are lifted at a mesoscale, flow over an orographic barrier and are prevented from flowing into deep valleys. To account for these effects, we used a variant of Eq. [3] and Eq. [4] with a linear search distance of 300 km in steps of 5° from 0° to 355° circular for each grid cell. The calculated leeward index was then scaled towards higher altitudes using:

$$E = H_L^{\frac{elev}{c}} \quad [5]$$

which rescales the strength of the exposition index relative to elevation (*elev*) from GMTED2010, given higher isolated valley higher isolations than lower ones. The correction constant *c* was set to 9000 m.

### Boundary layer correction:

Orographic precipitation effects are less pronounced just above the surface, as well as in the free atmosphere above the planetary boundary layer<sup>25,47,48</sup>. The highest impact is considered just at the boundary layer height (as indicator for the condensation level) where most of the cloud water content is located. While former studies used single ERA pressure levels, known to represent the main wind field patterns in a specific area<sup>24</sup>, the pressure level representing the prevailing wind directions at the boundary layer is usually not known a priori on a global basis. We therefore used the boundary layer height *B* as indicator of the pressure level that has the highest contribution to the wind effect. The boundary layer height has been interpolated using a B-spline interpolation to the CHELSA resolution. To recreated the typically nonlinear observed precipitation lapse rates<sup>49,50</sup> with terrain elevation, the wind effect grid *H* containing the windward (*H<sub>w</sub>*) and leeward (*H<sub>L</sub>*) index values was then proportionally distributed within an ERA grid cell the grid containing the boundary layer height *B* using:

$$H_{WB} = \frac{H_W}{1 - \left(\frac{|d| - d_{max}}{f}\right)} \quad [6]$$

$$H_{LB} = \frac{H_L}{1 - \left(\frac{|d| - d_{max}}{f}\right)} \quad [7]$$

with:

$$d = alt - B \quad [8]$$

and with *d* being the distance between a grid cell and the boundary layer height and *d<sub>max</sub>* being the maximum distance between the boundary layer height and all grid cell at the CHELSA resolution within an respective ERA grid cell, and *f* being a constant of 9000 m.

with:

$$B = B_{ERA} + elev_{ERA} + f \quad [9]$$

*B* being the height of the monthly means of daily mean boundary layer from ERA-Interim, *elev<sub>ERA</sub>* being the elevation of the ERA interim grid cell, and *f* being a constant of 500 m which takes into account that the level of highest precipitation is not necessarily at the lower bound of the boundary layer, but slightly higher<sup>47,48</sup>.

### Precipitation data from ERA-Interim

For accumulated parameters (total precipitation), we used the monthly means of daily forecast accumulations of total precipitation initialized at the synoptic hours 0:00 and 12:00. To calculate monthly precipitation sums, we added the synoptic monthly means at time 0:00, step 12 and time 12, step 12 and multiplied it by the number of days in in the respective month.

### Bias correction of ERA-Interim data using GPCC and GHCN data

Model-generated estimates of the surface precipitation are extracted from short range forecasts, which vary with forecast length. This drift in the short-range forecasts can be a problem for users of monthly and climatic means<sup>51</sup>. One very common approach is to calculate the difference

between baseline precipitation from the GCM and the observed precipitation and apply this ‘factor of change’ to historical observed time series to generate a synthetic time series<sup>52–54</sup>. We therefore performed three steps of bias correction.

### Monthly bias correction

We applied the monthly bias correction before the downscaling of the precipitation data on the ERA-interim precipitation values directly<sup>54</sup>. For this end, we used the monthly values of the gridded GPCC dataset<sup>13</sup> to calculate the monthly bias  $R_m$  caused by the spin up – spin down of the forecast algorithm<sup>51</sup> for each month from Jan. 1979 – Dec. 2013 using:

$$R_m = \frac{p_{GPCC}}{p_{ERA}} \quad [10]$$

We only used grid cells with climatic stations present for  $R_m$ . The spin up – spin down effect of ERA-Interim is spatially not independent, with a larger bias over high elevation terrain, or specific land forms such as tropical rainforests<sup>51</sup>. Based on this observation, we assumed that grid cells without stations share a similar bias as their neighbouring stations. To achieve a gap-free grid surface, we therefore interpolated the gaps in the  $R_m$  grid using a multilevel B-spline interpolation with 14 error levels. The gap-free bias correction surface  $R_m$  was then multiplied with the ERA-Interim precipitation to archive the bias corrected monthly precipitation sums  $p_m$ :

$$p_m = p_{ERA} \times R_m \quad [11]$$

### Monthly precipitation including orographic effects

To achieve the distribution of monthly precipitation sums  $p$  including orographic effects, we proportionally distributed the monthly bias corrected precipitation grids at the ERA resolution  $p_m$  onto the boundary layer corrected wind effect surface  $H$  using:

$$p = p_m * \frac{H}{\bar{H}} \quad [12]$$

where  $\bar{H}$  is the mean wind effect at ERA resolution.

### Mean annual precipitation sums

We calculated the mean annual precipitation sums as the mean annual sum of precipitation in the years 1979–2013. As slight errors in the precipitation sums can, however, accumulate over time, we applied two additional bias correction steps which accounts for biases at a small spatial scale and at a large spatial scale. Additional errors can be accumulated as the locations of the grid cell do not fully reflect the climatic dynamics of the underlying complex terrain<sup>55</sup>, as well as the location of the climatic stations.

### Large scale bias correction

A bias correction as applied for the monthly data would not work on the annual sums, as it would disregard the already included orographic wind effects at the finer resolution. The error also might change the modelled precipitation gradient in a specific area with differences in the maximum and minimum precipitation amounts (e.g. systematic error residuals of the spin up – spin down bias of the ERA-Interim forecast algorithm still present in the data). The applied B-spline interpolation of the monthly biases can also additionally create an error which might increase over time when precipitation sums are accumulated. We therefore opted for a bias correction of the maximum and minimum modelled precipitation values by the maximum and minimum precipitation values from the GPCC data. We therefore transferred the maximum and

minimum values from both CHELSA and GPCC to a 1.5° x 1.5° grid and applied a bias correction on the gradient using:

$$p_{cor} = p_{\min GPCC} + \left( \frac{p_{\max GPCC} - p_{\min GPCC}}{p_{\max} - p_{\min}} - p_{\min} \right) \quad [13]$$

where  $p_{\min}$  and  $p_{\max}$  are the maximum mean and minimum mean modelled precipitation values, and  $p_{\min GPCC}$  and  $p_{\max GPCC}$  are the respective GPCC maximum mean and minimum mean values. As the GPCC data is based on climatic stations, the distribution of stations is not necessarily even within a 1.5° x 1.5° grid cell. If the GPCC values only represent valley stations for example, high precipitation sums are most likely to be underestimated by GPCC. We therefore only used the GPCC maximum and minimum values in the GPCC grid where a) stations were available, and b) the minimum values were smaller than the modelled values, and the maximum values were larger than the modelled values.

### Small scale bias correction

Extreme values of precipitation at small spatial scales (e.g. single mountain tops) are often not caught by model algorithms, and are usually not reflected in the mean values at the ERA-Interim, or GPCC resolutions especially in complex terrain<sup>55</sup>. To include them, we used a small scale bias correction at the ERA-Interim resolution with data from climatic stations. For the GHCN station data we only used stations with full annual records in the years 1979-2013 and no obvious outliers or errors. The coordinates of the GHCN dataset are, however, of varying accuracy, and coordinates with only 1 decimal cannot accurately be placed within a specific grid cell at the CHELSA resolution. We therefore transferred the maximum and minimum annual precipitation sums recorded within an ERA-Interim grid cell for the modelled and observed precipitation values to the T255 grid. The selection for either GHCN maxima or minima compared to CHELSA maxima or minima was then done similarly to the large scale bias correction. The bias correction was performed using:

$$p_{cor2} = p_{\min GHCN} + \left( \frac{p_{\max GHCN} - p_{\min GHCN}}{p_{cor \max} - p_{cor \min}} - p_{cor \min} \right) \quad [14]$$

### Backward bias correction

To remove the bias responsible for the accumulated precipitation values from the monthly precipitation values  $p_m$ , we applied a backward bias correction using:

$$p_{mcor} = p_m \times \frac{\sum_{1979}^{2013} p_m}{35 p_{cor2}} \quad [15]$$

which removes the accumulated error from the monthly values and satisfies the condition that the sums of all month equals the corrected mean annual sums for 1979-2013  $p_{cor2}$ .

### Code availability

The codes used to calculate the CHELSA climatology are written in C++ and are included in SAGA Version 2.2.7, free available at [www.saga-gis.org](http://www.saga-gis.org) under the GNU public license including the necessary source codes. Calculations were done in SAGA Version 2.2.7 on the “Science Cloud” cloud computing facility of the University of Zurich [www.s3it.uzh.ch/infrastructure/sciencecloud/](http://www.s3it.uzh.ch/infrastructure/sciencecloud/).

### Data Records

The CHELSA data contains records for monthly mean temperature and precipitation values and derived annual mean values for the reference period 1979-2013 in form of SAGA grid files (Table 1). The files are freely available at [www.chelsa-climate.org](http://www.chelsa-climate.org).

The file format is SAGA. With the use of the associated .sdatt.aux.xml, the grid files can be opened in any GIS application including R. To open the files in R, we recommend the use of the function raster() from the raster package<sup>56</sup> with the .sdatt file extension.

## Technical Validation

To validate the results of the CHELSA algorithm, we compare it with different datasets that are available at comparable spatial and temporal resolution. A statistical comparison with different datasets is, however, complicated by the fact that most temperature and precipitation datasets are parameterized using similar observational data, leading to generally high correlations between climatic reanalyses.

### Large scale spatial comparison of precipitation patterns

To compare our precipitation data with those of other products, we first compared the spatial patterns of our results with those of the Tropical Rainfall Measuring Mission (TRMM)<sup>2,3</sup> combined multisatellite product TRMM/TMPA (3B43)<sup>57</sup> for the years 1998-2013. TRMM/TMPA (3B43)<sup>57</sup> is one of the few products that is available on a monthly basis on a high resolution and therefore allows comparison of the accumulated precipitation sums for a given reference period captured by CHELSA. A comparison of the two products shows high correlations for the mean annual sums ( $r=0.95$ ,  $\text{slope}=0.934$ ). Differences in the mean annual sums can mainly be detected in coastal regions, where CHELSA usually has higher rain sums compared to TRMM, and on the continents where CHELSA usually has lower values than TRMM/TMPA (3B43)<sup>57</sup> (Fig. 2). An explanation for these differences might be that TRMM/TMPA (3B43)<sup>57</sup> is not able to distinguish between windward and leeward sides of mountains below its  $0.25^\circ$  resolution and therefore under- or overestimates the rainfall amounts in these areas. This is especially visible on oceanic islands such as Hawai'i, where TRMM/TMPA (3B43)<sup>57</sup> is not able to capture the high amounts of rainfall<sup>58</sup> created by orographic effects and shows a severe dry bias compared to CHELSA. In general, the deviations between TRMM/TMPA (3B43)<sup>57</sup> and CHELSA are moderate, with only a few regions showing deviations above 500 mm/year in precipitation. For these regions, we would urge caution the use of the total annual sums of our model and suggest the use of multiple models from various sources.

### Small scale comparison of precipitation patterns

We conducted a small scale comparison of the precipitation patterns with three different models in the topographically and climatically highly complex terrain of Bhutan (Fig. 3). A comparison of the mean annual sums between TRMM/TMPA (3B43)<sup>57</sup>, Worldclim<sup>8</sup>, CHELSA, and the statistical downscaling approach of Böhner<sup>35</sup> shows similar patterns between all models at the mesoscale. The differences at the microscale are, however, severe between CHELSA and Böhner<sup>35</sup> compared to Worldclim<sup>8</sup>. There are only few climate stations in the region of Bhutan, which creates spurious correlations between elevation and precipitation in the ANUSPLIN algorithm of Worldclim. CHELSA and Böhner show a more consistent relation between the terrain features and the resulting precipitation patterns. A validation of these patterns is, however, complicated, as there are no independent climate stations available. Yet, a comparison with the patterns of cloud formations in this region<sup>59</sup> shows similarities in the patterns where clouds form and where higher precipitation amounts are predicted by CHELSA and Böhner (Fig. 3). Although the formation of clouds does not necessarily coincide with rainfall, there is generally a high correlation between the formation of clouds and the patterns of rainfall



especially in topographically complex terrain<sup>60</sup>. We therefore assume that our model is able to capture the topographic heterogeneity of precipitation at the small spatial scale rather well. We caution, however, towards the modelled absolute amounts of rainfall, which might be still under- or overestimated in certain cases compared to other models (Fig. 2).

### **Temporal correlations of precipitation and temperature patterns with comparable datasets**

We conducted the temporal comparison of CHELSA precipitation with other data with TRMM/TMPA (3B43)<sup>57</sup> on a monthly basis from 1998-2011 (Fig. 4). For the region covered by TRMM/TMPA (3B43)<sup>57</sup> ranging from 50°N to 50°S, correlations of the spatial distribution of precipitation are generally high (ranging from 0.80-0.90). There are also no systematic variations in the correlations (Fig. 4) during the analysed time period, indicating that the CHELSA algorithm performs spatially as well as temporally well for precipitation. Although station data was used to calibrate the CHELSA algorithm as well TRMM/TMPA (3B43) algorithm, we include a comparison between the station data, CHELSA, and TRMM/TMPA (3B43). This is, however, solely done to compare the patterns generated by CHELSA and TRMM/TMPA (3B43) relative to each other.

We compared CHELSA temperature data to that of MODIS (MOD11C3)<sup>61</sup> and GHCN Version 3<sup>11</sup>. Other high resolution products for temperature such as WorldClim do not have the same validation period as CHELSA. A comparison is therefore problematic due to the increase of global temperatures in the last decades. PRISM<sup>25</sup> is geographically restricted to the United States and therefore also not available for global comparisons. As climate station data from GHCN Version 3 is not directly used by the CHELSA algorithm for temperature, a comparison with station data is possible. As stations only provide limited information about the spatial patterns of temperature distribution, we additionally compared the CHELSA temperature patterns to those of MODIS (MOD11C3)<sup>61</sup> (Fig. 4). While MODIS (MOD11C3)<sup>61</sup> is able to detect the spatial patterns of temperature comparably well at a small spatial scale, it suffers from several drawbacks such as the inability to detect temperatures in regions with high cloud cover. Coefficients of determination between MODIS (MOD11C3)<sup>61</sup> and CHELSA temperatures range from 0.95 to 0.99 globally, between GHCN Version 3 and CHELSA temperatures range 0.96 to 0.99 globally, and between MODIS (MOD11C3) and GHCN Version 2 range from 0.83 – 0.97. Both CHELSA and MODIS show systematically lower correlations during the northern summer months which might indicate erroneous temperature values in the GHCN dataset. The high spatial correlation between CHELSA and MODIS (MOD11C3)<sup>61</sup> shows that CHELSA is able to predict that spatial patterns of temperature distributions well, and additionally accurately predicts the observed values of temperature on a small scale.

### **Acknowledgements**

The authors would like to acknowledge funding from the Swiss National Funds (SNF 147630, SNF 146906). We thank Sergio Maffioletti for implementing the CHELSA algorithm on the science cloud grid computing facility of the University of Zurich.

### **Author contributions**

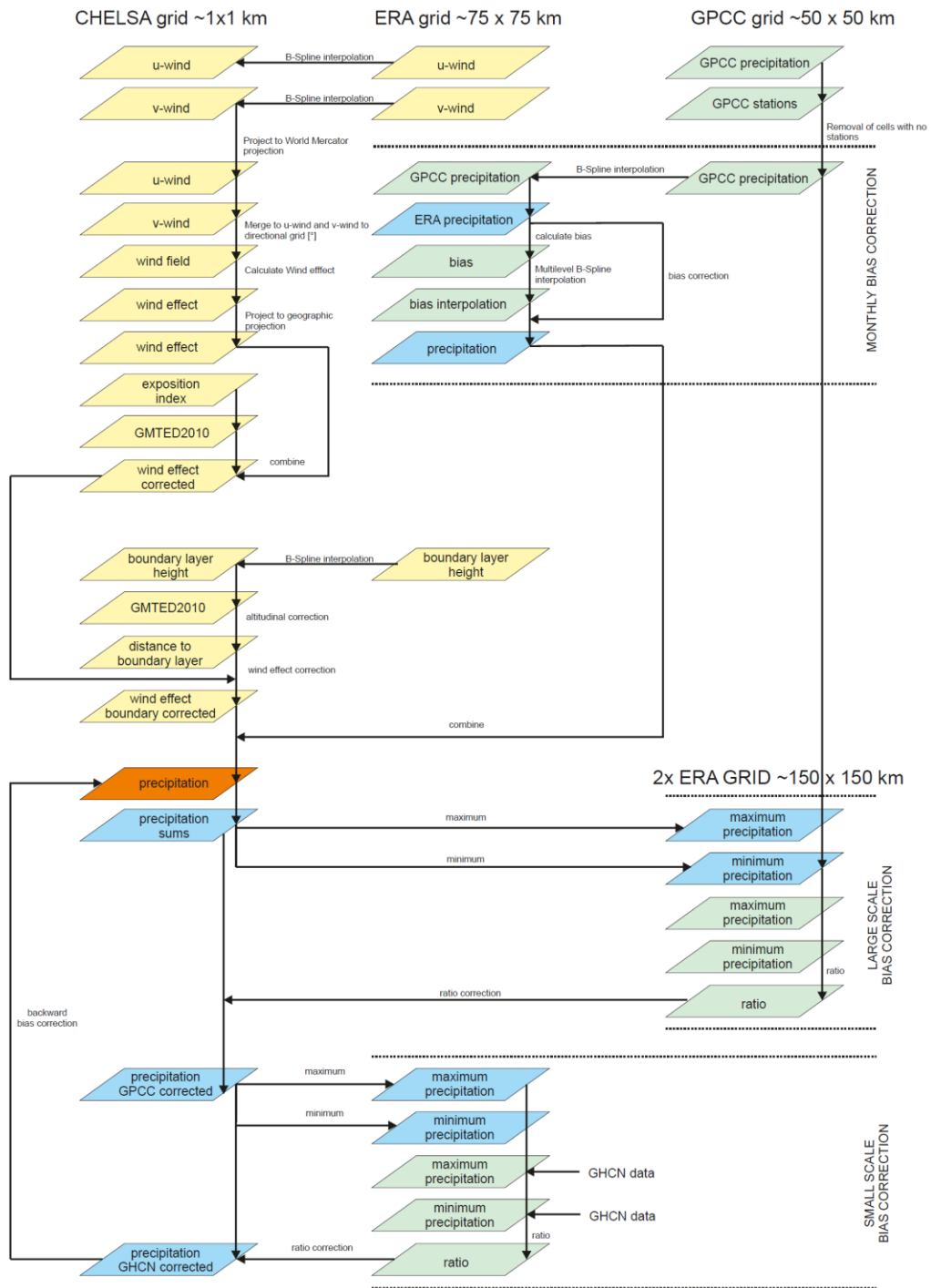
M.K. initiated the project. D.N.K., O.K., and T.K. developed the algorithms in close communication with J.B. R.W.S. compiled the GHCN data and removed the errors. M.K., H.K., P.L., and N.Z. provided the funding for the project. D.N.K. wrote the first draft of the manuscript and all authors contributed significantly to the revisions.

### **Competing interests**

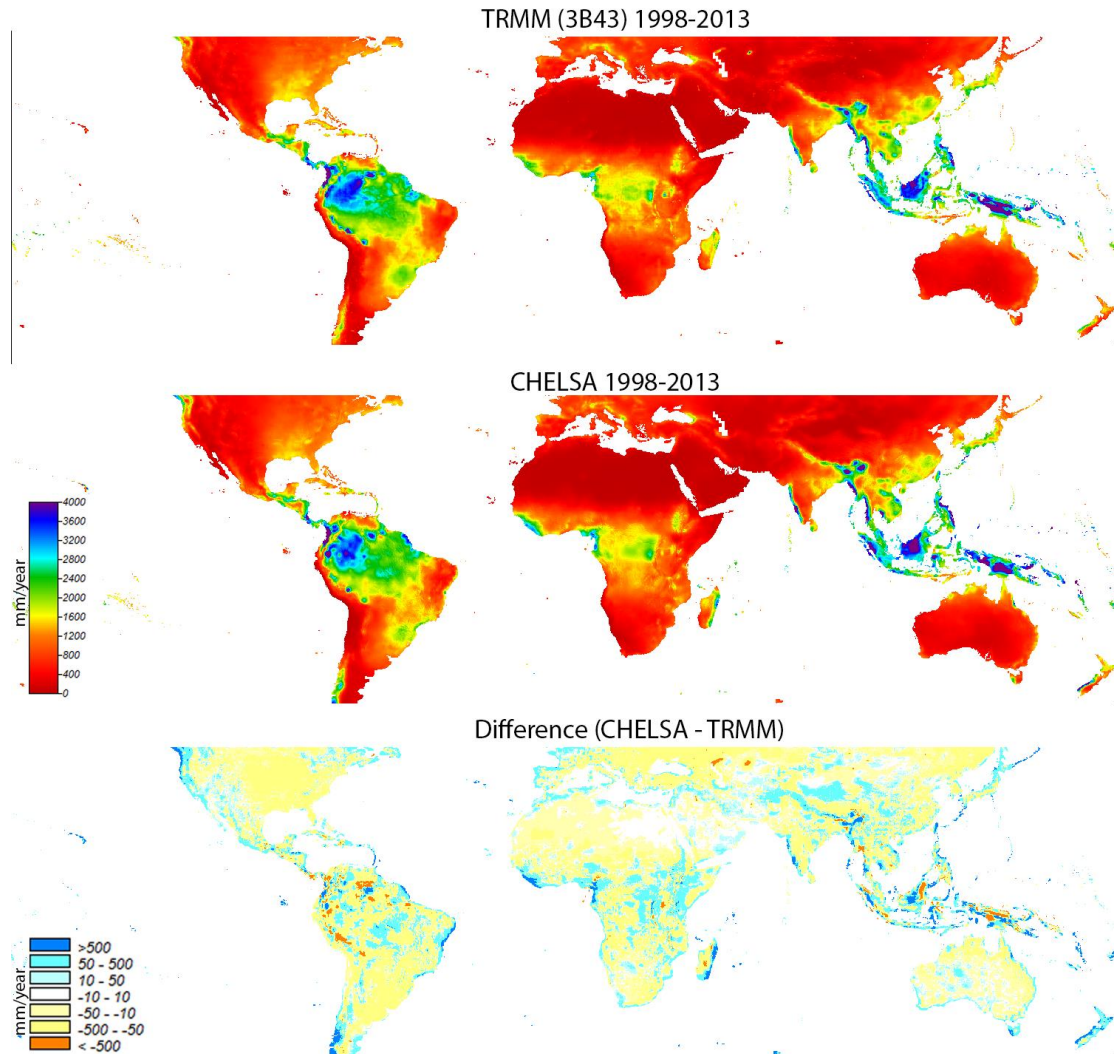
The authors declare no conflict of interest.

## Figures

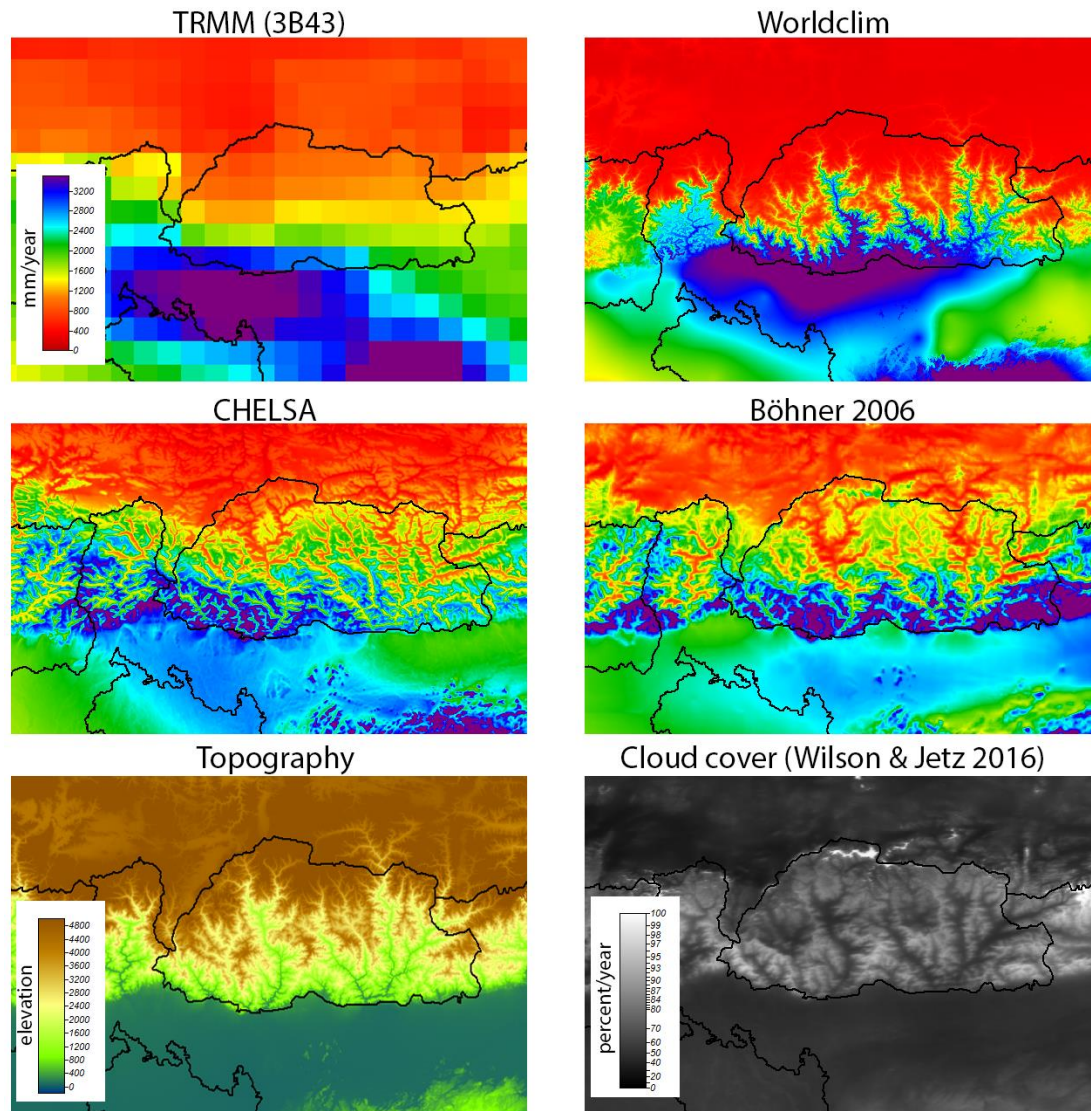
### THE CHELSA ALGORITHM



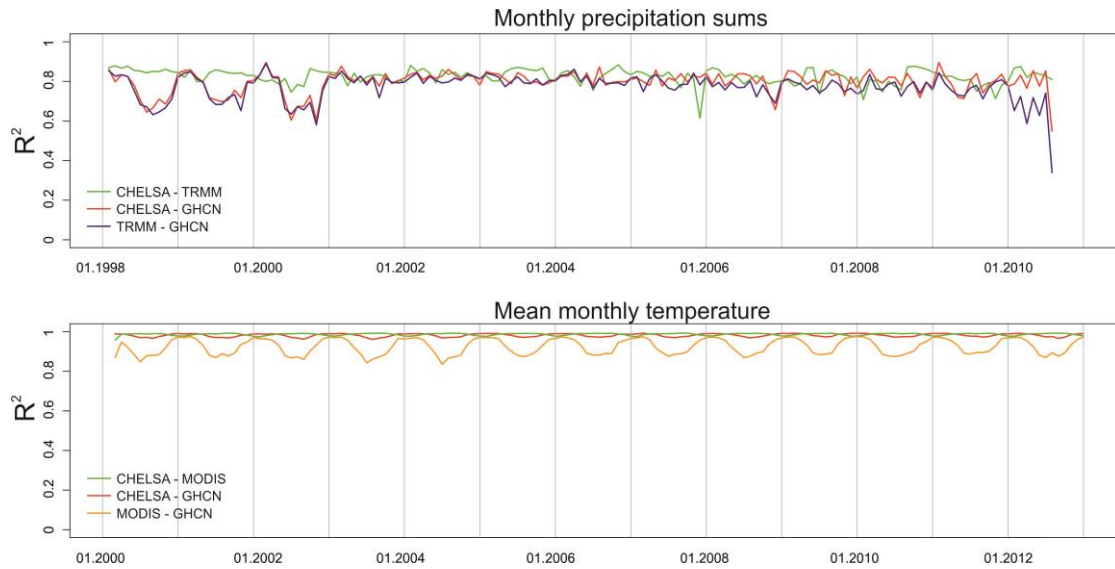
**Figure 1.** Workflow of the CHELSA algorithm for precipitation data. Resulting raster datasets (parallelograms) from each calculation step (arrows) are shown for each step of the algorithm. Predictor variables are indicated in yellow, raster datasets of the dependent variable (precipitation) are indicated in blue, and bias correction raster datasets are indicated in green. The monthly product from which the climatology is derived, and that is available for download as is indicated in orange.



**Figure 2.** Comparison of patterns of mean annual precipitation sums over the years 1998-2013 between TRMM (3B43) and CHELSA, and their respective differences on a global scale. CHELSA is generally drier over large land masses than TRMM and wetter along mountain ranges and coastal regions. An exception seems to be the Andes of Peru and Bolivia, and the high mountain regions of Papua New Guinea, where CHELSA shows lower amounts of precipitation.



**Figure 3.** Comparison of precipitation patterns in the complex terrain of Bhutan (country boundaries in black) between TRMM (3B43), Worldclim<sup>8</sup>, CHELSA, the statistical downscaling approach of Böhner<sup>35</sup>, the topography from GMTED2010, and the cloud cover climatology from Wilson & Jetz<sup>59</sup>. In this region, most precipitation falls during the SW-monsoon in the northern summer, when wet air masses from the SW are lifted at the south face of the Himalayas and dry until reaching the Tibetan high plateau. While the mesoscale patterns are in congruence between models, there are clear differences at the microscale. WorldClim predicts wet valleys and dry mountain faces, whereas CHELSA and Böhner<sup>35</sup> predict dry valleys and wet windward exposed mountain faces due to the inclusion of orographic predictors. CHELSA and Böhner<sup>35</sup> are also in closer congruence with the observed distribution of cloud in the area, which shows lower cloud cover in the isolated mountain valleys compared to the wind exposed mountain faces in the south.



**Figure 4.** Temporal comparison of the CHELSA algorithm with TRMM, GHCN, and MODIS. Coefficients of determination give the global correlation between products for a specific month. CHELSA precipitations shows significantly higher correlations with GHCN (Wilcoxon Test:  $W=8370$ ,  $P<0.001$ ), and CHELSA temperatures show significantly higher correlations with GHCN (Wilcoxon Test:  $W=23254$ ,  $P<0.001$ ). Correlations between CHELSA and TRMM are generally high for precipitation (Mean  $R^2=0.82$ ), and high between CHELSA and MODIS for temperature (Mean  $R^2=0.99$ ).



## References

1. Funk, C. *et al.* The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* **2**, 150066 (2015).
2. Biasutti, M., Yuter, S. E., Burleyson, C. D. & Sobel, A. H. Very high resolution rainfall patterns measured by TRMM precipitation radar: seasonal and diurnal cycles. *Clim. Dyn.* **39**, 239–258 (2011).
3. Huffman, G. J. *et al.* The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *J. Hydrometeorol.* **8**, 38–55 (2007).
4. Maraun, D. *et al.* Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.* **48**, RG3003 (2010).
5. Wood, A. W., Leung, L. R., Sridhar, V. & Lettenmaier, D. P. Hydrologic Implications of Dynamical and Statistical Approaches to Downscaling Climate Model Outputs. *Clim. Change* **62**, 189–216 (2004).
6. Wilby, R. L. *et al.* Statistical downscaling of general circulation model output: A comparison of methods. *Water Resour. Res.* **34**, 2995–3008 (1998).
7. Schmidli, J., Frei, C. & Vidale, P. L. Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *Int. J. Climatol.* **26**, 679–689 (2006).
8. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
9. Schoof, J. t. & Pryor, S. c. Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. *Int. J. Climatol.* **21**, 773–790 (2001).
10. Soria-Auza, R. W. *et al.* Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. *Ecol. Model.* **221**, 1221–1229 (2010).
11. Lawrimore, J. H. *et al.* An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J. Geophys. Res. Atmospheres* **116**, (2011).
12. Peterson, T. C. & Vose, R. S. An overview of the Global Historical Climatology Network temperature database. *Bull. Am. Meteorol. Soc.* **78**, 2837–2849 (1997).
13. Schneider, U. *et al.* GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theor. Appl. Climatol.* **115**, 15–40 (2013).
14. Becker, A. *et al.* A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst Sci Data* **5**, 71–99 (2013).
15. New, M., Hulme, M. & Jones, P. Representing Twentieth-Century Space–Time Climate Variability. Part I: Development of a 1961–90 Mean Monthly Terrestrial Climatology. *J. Clim.* **12**, 829–856 (1999).
16. Harris, I., Jones, P. d., Osborn, T. j. & Lister, D. h. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
17. Kalnay, E. *et al.* The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
18. Dee, D. P. *et al.* The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
19. Lloyd, C. D. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *J. Hydrol.* **308**, 128–150 (2005).
20. Jarvis, C. H. & Stuart, N. A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part II: The interaction between number of guiding variables and the type of interpolation method. *J. Appl. Meteorol.* **40**, 1075–1084 (2001).
21. Pebesma, E. J. The Role of External Variables and GIS Databases in Geostatistical Analysis. *Trans. GIS* **10**, 615–632 (2006).

22. Böhner, J. & Antonic, O. in *T. Hengl, & H. I. Reuter (eds.) Geomorphometry: Concepts, Software, Applications* 195–226 (Elsevier Science, 2009).
23. Wilby, R. L. & Wigley, T. M. L. Downscaling general circulation model output: a review of methods and limitations. *Prog. Phys. Geogr.* **21**, 530–548 (1997).
24. Gerlitz, L., Conrad, O. & Böhner, J. Large-scale atmospheric forcing and topographic modification of precipitation rates over High Asia – a neural-network-based approach. *Earth Syst Dynam* **6**, 61–81 (2015).
25. Daly, C., Taylor, G. H. & Gibson, W. P. The PRISM approach to mapping precipitation and temperature. in *Proc., 10th AMS Conf. on Applied Climatology* 20–23 (Citeseer, 1997).
26. Berrisford, P. *et al.* The ERA-interim archive. *ERA Rep. Ser.* 1–16 (2009).
27. Berrisford, P. *et al.* Atmospheric conservation properties in ERA-Interim. *Q. J. R. Meteorol. Soc.* **137**, 1381–1399 (2011).
28. Gao, L. *et al.* Statistical Downscaling of ERA-Interim Forecast Precipitation Data in Complex Terrain Using LASSO Algorithm, Statistical Downscaling of ERA-Interim Forecast Precipitation Data in Complex Terrain Using LASSO Algorithm. *Adv. Meteorol. Adv. Meteorol.* **2014**, **2014**, e472741 (2014).
29. Bao, X. & Zhang, F. Evaluation of NCEP–CFSR, NCEP–NCAR, ERA-Interim, and ERA-40 Reanalysis Datasets against Independent Sounding Observations over the Tibetan Plateau. *J. Clim.* **26**, 206–214 (2012).
30. Betts, A. K., Köhler, M. & Zhang, Y. Comparison of river basin hydrometeorology in ERA-Interim and ERA-40 reanalyses with observations. *J. Geophys. Res. Atmospheres* **114**, D02101 (2009).
31. Hansen, J., Sato, M. & Ruedy, R. Radiative forcing and climate response. *J. Geophys. Res. Atmospheres* **102**, 6831–6864 (1997).
32. Rolland, C. Spatial and Seasonal Variations of Air Temperature Lapse Rates in Alpine Regions. *J. Clim.* **16**, 1032–1046 (2003).
33. Minder, J. R., Mote, P. W. & Lundquist, J. D. Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains. *J. Geophys. Res. Atmospheres* **115**, D14122 (2010).
34. Danielson, J. J. & Gesch, D. B. *Global multi-resolution terrain elevation data 2010 (GMTED2010)*. (US Geological Survey, 2011).
35. Böhner, J. General climatic controls and topoclimatic variations in Central and High Asia. *Boreas* **35**, 279–295 (2006).
36. Spreen, W. C. A determination of the effect of topography upon precipitation. *Eos Trans. Am. Geophys. Union* **28**, 285–290 (1947).
37. Gao, X., Xu, Y., Zhao, Z., Pal, J. S. & Giorgi, F. On the role of resolution and topography in the simulation of East Asia precipitation. *Theor. Appl. Climatol.* **86**, 173–185 (2006).
38. Basist, A., Bell, G. D. & Meentemeyer, V. Statistical Relationships between Topography and Precipitation Patterns. *J. Clim.* **7**, 1305–1315 (1994).
39. Daly, C., Neilson, R. P. & Phillips, D. L. A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. *J. Appl. Meteorol.* **33**, 140–158 (1994).
40. Sevruck, B. in *Climatic Change at High Elevation Sites* (eds. Diaz, H. F., Beniston, M. & Bradley, R. S.) 123–137 (Springer Netherlands, 1997).
41. Rotunno, R. & Houze, R. A. Lessons on orographic precipitation from the Mesoscale Alpine Programme. *Q. J. R. Meteorol. Soc.* **133**, 811–830 (2007).
42. Weischet, W. & Endlicher, W. Einführung in die allgemeine Klimatologie. (2008).
43. Roe, G. H. Orographic Precipitation. *Annu. Rev. Earth Planet. Sci.* **33**, 645–671 (2005).
44. Colle, B. A. Sensitivity of Orographic Precipitation to Changing Ambient Conditions and Terrain Geometries: An Idealized Modeling Perspective. *J. Atmospheric Sci.* **61**, 588–606 (2004).
45. Sinclair, M. R. A Diagnostic Model for Estimating Orographic Precipitation. *J. Appl. Meteorol.* **33**, 1163–1175 (1994).
46. Smith, R. B. & Barstad, I. A Linear Theory of Orographic Precipitation. *J. Atmospheric Sci.* **61**, 1377–1391 (2004).

47. Oke, T. R. *Boundary layer climates*. (Routledge, 2002).
48. Stull, R. B. *An introduction to boundary layer meteorology*. **13**, (Springer Science & Business Media, 2012).
49. Körner, C. The use of 'altitude' in ecological research. *Trends Ecol. Evol.* **22**, 569–574 (2007).
50. Alam, F. C. K. Distribution of precipitation in mountainous areas of West Pakistan. *Distrib. Precip. Mt. AREAS Vol. II* (1973).
51. Kållberg, P. *Forecast drift in ERA-Interim*. (European Centre for Medium Range Weather Forecasts, 2011).
52. Lafon, T., Dadson, S., Buys, G. & Prudhomme, C. Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods. *Int. J. Climatol.* **33**, 1367–1381 (2013).
53. Arnell, N. W., Hudson, D. A. & Jones, R. G. Climate change scenarios from a regional climate model: Estimating change in runoff in southern Africa. *J. Geophys. Res. Atmospheres* **108**, 4519 (2003).
54. Molteni, F. A 'historical' approach to the rescaling of ERA-Interim precipitation, internal technical note. (European Centre for Medium Range Weather Forecasts, 2013).
55. Maraun, D. & Widmann, M. The representation of location by a regional climate model in complex terrain. *Hydrol Earth Syst Sci* **19**, 3449–3456 (2015).
56. Hijmans, R. J. & van Etten, J. raster: Geographic data analysis and modeling. *R Package Version 2*, 15 (2014).
57. Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC). *TRMM/TMPA 3B43 TRMM and Other Sources Monthly Rainfall Product V7*. (2011).
58. Giambelluca, T. W. *et al.* Online rainfall atlas of Hawai'i. *Bull. Am. Meteorol. Soc.* **94**, 313–316 (2013).
59. Wilson, A. M. & Jetz, W. Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. *PLOS Biol* **14**, e1002415 (2016).
60. Pruppacher, H. R., Klett, J. D. & Wang, P. K. Microphysics of clouds and precipitation. (1998).
61. NASA LP DAAC. *MODIS/Terra Land Surface Temperature and Emissivity Monthly L3 Global 0.05Deg CMG*. (NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, 2015).



Table 1. Data records.

Filename	variable	unit	timespan	format	bias correction	remarks
CHELSA_temp_month_year.sgrd	mean monthly temperature	°C	month	SAGA	none	File containing the grid information
CHELSA_temp_month_year.sdat	mean monthly temperature	°C	month	SAGA	none	File containing the data values of the grid
CHELSA_temp_month_year.prj	mean monthly temperature	°C	month	SAGA	none	File containing the projection parameters
CHELSA_temp_month_year.sdat.aux.xml	mean monthly temperature	°C	month	SAGA	none	File containing auxiliary information
CHELSA_prec_month_year.sgrd	mean monthly precipitation sums	mm/month	month	SAGA	none	File containing the grid information
CHELSA_prec_month_year.sdat	mean monthly precipitation sums	mm/month	month	SAGA	none	File containing the data values of the grid
CHELSA_prec_month_year.prj	mean monthly precipitation sums	mm/month	month	SAGA	none	File containing the projection parameters
CHELSA_prec_month_year.sdat.aux.xml	mean monthly precipitation sums	mm/month	month	SAGA	none	File containing auxiliary information
CHELSA_prec_GHCN_month_year.sgrd	mean monthly precipitation sums	mm/month	month	SAGA	GHCN	File containing the grid information
CHELSA_prec_GHCN_month_year.sdat	mean monthly precipitation sums	mm/month	month	SAGA	GHCN	File containing the data values of the grid
CHELSA_prec_GHCN_month_year.prj	mean monthly precipitation sums	mm/month	month	SAGA	GHCN	File containing the projection parameters
CHELSA_prec_GHCN_month_year.sdat.aux.xml	mean monthly precipitation sums	mm/month	month 1979-	SAGA	GHCN	File containing auxiliary information
CHELSA_prec_GHCN_1979-2013.sgrd	mean annual precipitation sums	mm/year	2013 1979-	SAGA	GHCN	File containing the grid information
CHELSA_prec_GHCN_1979-2013.sdat	mean annual precipitation sums	mm/year	2013 1979-	SAGA	GHCN	File containing the data values of the grid
CHELSA_prec_GHCN_1979-2013.prj	mean annual precipitation sums	mm/year	2013 1979-	SAGA	GHCN	File containing the projection parameters
CHELSA_prec_GHCN_1979-2013.sdat.aux.xml	mean annual precipitation sums	mm/year	2013 1979-	SAGA	GHCN	File containing auxiliary information
CHELSA_temp_1979-2013.sgrd	mean annual temperature	°C	2013 1979-	SAGA	none	File containing the grid information
CHELSA_temp_1979-2013.sdat	mean annual temperature	°C	2013 1979-	SAGA	none	File containing the data values of the grid
CHELSA_temp_1979-2013.prj	mean annual temperature	°C	2013 1979-	SAGA	none	File containing the projection parameters
CHELSA_temp_1979-2013.sdat.aux.xml	mean annual temperature	°C	2013 1979-	SAGA	none	File containing auxiliary information
CHELSA_prec_GHCN_month_1979-2013.sgrd	mean monthly precipitation sums	mm/year	2013 1979-	SAGA	GHCN	File containing the grid information
CHELSA_prec_GHCN_month_1979-2013.sdat	mean monthly precipitation sums	mm/year	2013 1979-	SAGA	GHCN	File containing the data values of the grid
CHELSA_prec_GHCN_month_1979-2013.prj	mean monthly precipitation sums	mm/year	2013	SAGA	GHCN	File containing the projection parameters

CHELSA_prec_GHCN_month_1979-2013.sdat.aux.xml	mean monthly precipitation sums	mm/year	1979-2013	SAGA	GHCN	File containing auxiliary information
CHELSA_temp_month_1979-2013.sgrd	mean monthly temperature	°C	1979-2013	SAGA	none	File containing the grid information
CHELSA_temp_month_1979-2013.sdat	mean monthly temperature	°C	1979-2013	SAGA	none	File containing the data values of the grid
CHELSA_temp_month_1979-2013.prj	mean monthly temperature	°C	1979-2013	SAGA	none	File containing the projection parameters
CHELSA_temp_month_1979-2013.sdat.aux.xml	mean monthly temperature	°C	1979-2013	SAGA	none	File containing auxiliary information

